Signal Inpainting from Fourier Magnitudes

Louis Bahrman[1,2]
[1]*LTCI, Télécom Paris, Institut Polytechnique de Paris*
Paris, France
louis.bahrman@telecom-paris.fr

Marina Krémé[2], Paul Magron[2], Antoine Deleforge[2]
[2]*Université de Lorraine, CNRS, Inria, LORIA*
Nancy, France
{ama-marina.kreme, paul.magron, antoine.deleforge}@inria.fr

***Abstract*—Signal inpainting is the task of restoring degraded or missing samples in a signal. In this paper we address signal inpainting when discrete Fourier magnitudes are observed. We propose a mathematical formulation of the problem that highlights its connection with phase retrieval, and we introduce two methods for solving it. First, we derive an alternating minimization scheme, which shares similarities with the Gerchberg-Saxton algorithm, a classical phase retrieval method. Second, we propose a convex relaxation of the problem, which is inspired by recent approaches that reformulate phase retrieval into a semi-definite program. We assess the potential of these methods for the task of inpainting gaps in speech signals. Our methods exhibit both a high probability of recovering the original signals and robustness to magnitude noise.**

***Index Terms*—Signal inpainting, phase retrieval, audio restoration, convex relaxation, alternating minimization.**

I. Introduction

Signal inpainting [1] is an inverse problem that consists in restoring signals degraded by sample loss. Such a problem typically arises as a result of degradation during signal transmission (packet loss concealment [2]) or in digitization of physically degraded media. Inpainting can also be used to restore signal samples subject to a degradation so heavy that the information about the samples can be considered lost (e.g., signal clipping [3] or impulsive noises [4]). More specifically, in this paper we focus on inpainting compact gaps, which occurs, e.g., when an audio signal is corrupted with clicks [5].

Approaches that tackle this issue can be divided into two categories depending on the number of missing samples or duration of the gaps. When considering short gaps (less than 100 ms), approaches based on autoregressive modeling [6], convex optimization [7], sparse modeling [8], or Bayesian estimation [9] have shown promising results. Conversely, approaches based on sinusoidal modeling [10] or graphs [11] are more suitable for longer gaps (more than 100 ms). However, signal inpainting remains a challenging problem, and these approaches are not adapted to scenarios where some additional information about the signal is available.

We focus here on a setting where the missing gap is contained in a short context window, while the *discrete Fourier magnitudes* of the complete signal on that segment are observed. A similar magnitude-informed setting was notably studied in [12] and [13] in the different context of source separation. Beyond the fact that this problem is open and has not received specific attention in the literature, the motivation behind studying it is the existence of a vast literature dedicated to modeling and processing short-term Fourier magnitudes in the audio literature, e.g., using nonnegative matrix factorization [14] or more recently variational auto-encoders [15]. This is because the Fourier magnitudes of natural signals tend to exhibit smoother and hence more predictable evolution than their respective Fourier phases. In this work, we assume that such a magnitude model has been leveraged beforehand, and we focus on the latest part of the inpainting problem, thus considering that some magnitude estimates are available, possibly up to errors.

The problem then shares a close connection with phase retrieval [16], the task of retrieving a signal from nonnegative measurements (usually magnitudes of a set of inner products). From the seminal early works of Gerchberg and Saxton [17], this task has been revived over the last decade with the development of novel optimization approaches based on gradient descent [18] or convex relaxations [19], [20]. While phase retrieval has experienced considerable progress in recent years, its connection to signal inpainting is largely left to explore, from the theoretical, methodological and applicative standpoints.

This papers aims at contributing to bridging this gap by formulating signal inpainting from Fourier magnitudes as a constrained phase retrieval problem. Inspired by phase retrieval algorithms, we derive two methods to tackle it, based on alternating minimization (AM) and convex relaxation (CR). Experiments conducted on speech signals reveal the potential of these techniques for audio inpainting, since they exhibit a high probability of recovering the original signal, as well as some robustness to magnitude errors. The approach compares favorably to the sate-of-the-art sparsity-based inpainting method in [8] as long as magnitudes are observed with sufficient accuracy.

The rest of this paper is organized as follows. Section II formulates the problem and introduces two methods for solving it. Experiments on speech signals are conducted in Section III. Finally, Section IV draws some concluding remarks.

II. Methods

A. From inpainting to phase retrieval

Let $\boldsymbol{x}^\natural \in \mathbb{R}^L$ denote a signal. We partition its support $\{0, \dots, L-1\}$ into two sets v and $\bar{v}$ such that $\boldsymbol{x}^\natural_v \in \mathbb{R}^{L-d}$

This work was made with the support of the French National Research Agency through project DENISE (ANR-20-CE48-0013), and was conducted while L. Bahrman was an intern with Inria in Nancy, France.

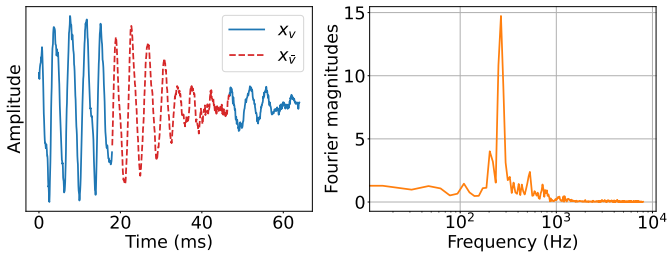

Fig. 1. Illustration of restoring the missing samples of a signal (left), assuming its Fourier magnitudes (right) have been estimated beforehand.

and $\boldsymbol{x}_{\bar{v}}^{\natural} \in \mathbb{R}^d$ denote the observed and missing samples, respectively, and where $d \leq L$ denotes the number of missing samples, whose location $\bar{v}$ is assumed to be known. Note that the missing samples $\mathbf{x}_{\bar{v}}$ can be non-contiguous. We also assume that the magnitudes of its discrete Fourier transform (DFT) $\boldsymbol{b} \in \mathbb{R}_+^L$ are observed. We consider a *complete* DFT (that is, it does not use zero-padding), whose matrix is denoted $\boldsymbol{\Phi} \in \mathbb{C}^{L \times L}$. The task of signal inpainting from Fourier magnitudes is illustrated in Fig. 1 and can be stated as:

$$\text{Find} \quad \boldsymbol{x} \in \mathbb{R}^L \quad \text{such that } |\boldsymbol{\Phi}\boldsymbol{x}| = \boldsymbol{b} \text{ and } \boldsymbol{x}_v = \boldsymbol{x}_v^{\natural}. \tag{1}$$

Let us formulate (1) as the following optimization problem:

$$\min_{\boldsymbol{x} \in \mathbb{R}^L} \| |\boldsymbol{\Phi}\boldsymbol{x}| - \boldsymbol{b} \|^2 \quad \text{s. t.} \quad \boldsymbol{x}_v = \boldsymbol{x}_v^{\natural}, \tag{2}$$

where $\|.\|$ denotes the Euclidean norm[1]. Problem (2) reads as a constrained phase retrieval problem. If the whole signal is missing ($d = L$), the constraint vanishes and it reduces to Fourier phase retrieval. We theoretically show in a supplementary material that for almost all $\boldsymbol{x}^{\natural} \in \mathbb{R}^L$, this problem admits a *unique solution* if at most 33% of the signal is missing [23]. Note that Fourier phase retrieval has been extensively investigated from a theoretical perspective; we refer the interested reader to the book chapter [24] for a comprehensive review on uniqueness guarantees. The cases where one sample or half of the samples at the end of the signal are observed were respectively covered some time ago in [25] and in [26]. However, both studies considered DFT with zero-padding. Our supplemental theorem complements these results by treating the more challenging unpadded DFT case, using a different proof technique.

Notwithstanding our uniqueness result, the problem is non-convex and can be viewed as an instance of quadratic programming, which is known to be NP-hard in general [27]. As pointed out in the review [24], *"there is* [currently] *no algorithm that knows how to exploit the given entries to recover the complete signal in a stable and efficient manner"*. This highlights the need for efficient algorithms that provide a solution. The cornerstone of our approaches lies in introducing an auxiliary phase variable $\boldsymbol{u} \in \mathbb{C}^L$ such that $|\boldsymbol{u}| = 1$. Then, $\boldsymbol{b} = |\operatorname{diag}(\boldsymbol{b})\boldsymbol{u}|$, where $\operatorname{diag}(\boldsymbol{b})$ is the diagonal matrix whose entries are given by the vector $\boldsymbol{b}$. We then turn our attention to minimizing the following auxiliary function, which is *exact* with respect to (2) in the sense of [28]:

$$\min_{\boldsymbol{x} \in \mathbb{R}^L, \boldsymbol{u} \in \mathbb{C}^L} \|\boldsymbol{\Phi}\boldsymbol{x} - \operatorname{diag}(\boldsymbol{b})\boldsymbol{u}\|^2 \text{ s.t. } \boldsymbol{x}_v = \boldsymbol{x}_v^{\natural}, \ |\boldsymbol{u}| = 1. \tag{3}$$

B. Alternating minimization

The first approach we propose to solve (3) is an AM scheme. Let us first fix $\boldsymbol{u}$ and derive the update for $\boldsymbol{x}$, for which we propose to incorporate the constraint $\boldsymbol{x}_v = \boldsymbol{x}_v^{\natural}$ within the loss function. To that end, let us reorder $\boldsymbol{x}$ as $\boldsymbol{x} = \begin{bmatrix} \boldsymbol{x}_{\bar{v}} \\ \boldsymbol{x}_v \end{bmatrix}$ (and similarly for $\boldsymbol{x}^{\natural}$), and split $\boldsymbol{\Phi}$ accordingly as $\boldsymbol{\Phi} = [\boldsymbol{\Phi}_{\bar{v}}, \boldsymbol{\Phi}_v]$, with $\boldsymbol{\Phi}_{\bar{v}} \in \mathbb{C}^{L \times d}$ and $\boldsymbol{\Phi}_v \in \mathbb{C}^{L \times (L-d)}$. Using these notations, we have $\boldsymbol{\Phi}\boldsymbol{x} = \boldsymbol{\Phi}_{\bar{v}}\boldsymbol{x}_{\bar{v}} + \boldsymbol{\Phi}_v\boldsymbol{x}_v$, and (3) rewrites:

$$\min_{\boldsymbol{x}_{\bar{v}} \in \mathbb{R}^d} \left\| \boldsymbol{\Phi}_{\bar{v}}\boldsymbol{x}_{\bar{v}} + \boldsymbol{\Phi}_v\boldsymbol{x}_v^{\natural} - \operatorname{diag}(\boldsymbol{b})\boldsymbol{u} \right\|^2. \tag{4}$$

Since $\boldsymbol{\Phi}_{\bar{v}}$ is full-rank, it has a left inverse which is its Hermitian transpose $\boldsymbol{\Phi}_{\bar{v}}^{\mathsf{H}}$. Besides, recall that since the DFT is an orthogonal transform, then $\boldsymbol{\Phi}_{\bar{v}}^{\mathsf{H}}\boldsymbol{\Phi}_v = 0$. Altogether this yields the following solution to (4):

$$\boldsymbol{x}_{\bar{v}} = \boldsymbol{\Phi}_{\bar{v}}^{\mathsf{H}} \operatorname{diag}(\boldsymbol{b})\boldsymbol{u}. \tag{5}$$

Let us now derive the update for $\boldsymbol{u}$ when $\boldsymbol{x}$ is fixed, for which (3) rewrites:

$$\min_{\boldsymbol{u} \in \mathbb{C}^L} \|\boldsymbol{\Phi}\boldsymbol{x} - \operatorname{diag}(\boldsymbol{b})\boldsymbol{u}\|^2 \quad \text{s. t.} \quad |\boldsymbol{u}| = 1, \tag{6}$$

which is straightforward to solve:

$$\boldsymbol{u} = \frac{\boldsymbol{\Phi}\boldsymbol{x}}{|\boldsymbol{\Phi}\boldsymbol{x}|}. \tag{7}$$

Alternating (5) and (7) yields a procedure summarized in Algorithm 1. We discuss the initialization strategy in Section III-A. Note that at line 5 we apply the real part function to ensure a real-valued signal estimate[2].

Remark: Algorithm 1 consists in computing the DFT of a signal, setting its magnitude to a target value, inverting the DFT, putting back the observed samples, and repeating. As such, it is similar to the Gerchberg-Saxton algorithm [17] except that the signal-domain magnitude constraint is here replaced with a projection onto the partially observed samples. Besides, Algorithm 1 can be seen as a member of the general family of *constrained phase retrieval* algorithms reviewed in [24]. While such algorithms are generally heuristically derived, the proposed alternate minimization scheme on an exact auxiliary function is guaranteed to converge - though not necessarily to a global minimum - by construction.

C. Convex relaxation

Let us now derive a method inspired by the PhaseCut algorithm [20], which is based on a CR of the problem. This approach consists in reformulating phase retrieval as a constrained trace minimization problem by lifting it to a higher dimensional space and relaxing the rank-one constraint. The

[1]Recent works such as [21], [22] have investigated alternative discrepancy measures for phase retrieval. We focus on the Euclidean norm in this study.

[2]It can be proven rigorously that doing so does not affect convergence guarantees as in [22], but we do not detail this here for brevity.

Algorithm 1 AM for signal inpainting

Input: $\begin{cases} \boldsymbol{b} \in \mathbb{R}_+^L : \text{observations}, \ \boldsymbol{x}_v^\natural : \text{known signal} \\ \boldsymbol{\Phi} : \text{Fourier matrix}, \end{cases}$

Initialize $\boldsymbol{x}_{\bar{v}}^{(0)}$ and $\boldsymbol{x}^{(0)} \leftarrow \begin{bmatrix} \boldsymbol{x}_{\bar{v}}^{(0)} \\ \boldsymbol{x}_v^\natural \end{bmatrix}$
$i \leftarrow 0$
while convergence not reached **do**
  $\boldsymbol{u}^{(i+1)} \leftarrow \frac{\boldsymbol{\Phi}\boldsymbol{x}^{(i)}}{|\boldsymbol{\Phi}\boldsymbol{x}^{(i)}|}$
  $\boldsymbol{x}_{\bar{v}}^{(i+1)} \leftarrow \Re\left(\boldsymbol{\Phi}_{\bar{v}}^{\mathsf{H}} \operatorname{diag}(\boldsymbol{b})\boldsymbol{u}^{(i+1)}\right)$
  $\boldsymbol{x}^{(i+1)} \leftarrow \begin{bmatrix} \boldsymbol{x}_{\bar{v}}^{(i+1)} \\ \boldsymbol{x}_v^\natural \end{bmatrix}$
  $i \leftarrow i+1$
end while

Output: Reconstructed signal $\boldsymbol{x}^{(i)}$

resulting problem can then be efficiently solved via semi-definite programming. We consider the formulation (3) in which we inject the expression of $\boldsymbol{x}$ given by (5):

$$\min_{\boldsymbol{u}\in\mathbb{C}^L} \left\|(\boldsymbol{\Phi}_{\bar{v}}\boldsymbol{\Phi}_{\bar{v}}^{\mathsf{H}} - \boldsymbol{I}) \operatorname{diag}(\boldsymbol{b})\boldsymbol{u} + \boldsymbol{\Phi}_v \boldsymbol{x}_v\right\|^2 \text{ s.t. } |\boldsymbol{u}| = 1. \quad (8)$$

Now, let us introduce the following:

$$\widetilde{\boldsymbol{m}} := [(\boldsymbol{\Phi}_{\bar{v}}\boldsymbol{\Phi}_{\bar{v}}^{\mathsf{H}} - \boldsymbol{I}) \operatorname{diag}(\boldsymbol{b}), \boldsymbol{\Phi}_v \boldsymbol{x}_v] \text{ and } \widetilde{\boldsymbol{u}} = \begin{bmatrix} \boldsymbol{u} \\ 1 \end{bmatrix}, \quad (9)$$

from which we can rewrite (8) as:

$$\min_{\widetilde{\boldsymbol{u}}\in\mathbb{C}^{L+1}} \|\widetilde{\boldsymbol{m}}\widetilde{\boldsymbol{u}}\|^2 \quad \text{s.t.} \quad |\widetilde{\boldsymbol{u}}| = 1 \text{ and } \widetilde{\boldsymbol{u}}[L] = 1. \quad (10)$$

Drawing on [20], we lift and relax (10) to the following convex problem:

$$\min_{\widetilde{\boldsymbol{U}}\in\mathbb{C}^{(L+1)\times(L+1)}} \operatorname{trace}(\widetilde{\boldsymbol{M}}\widetilde{\boldsymbol{U}}) \quad \text{s.t.} \quad \operatorname{diag}(\widetilde{\boldsymbol{U}}) = \mathbf{1}, \ \widetilde{\boldsymbol{U}} \succeq 0, \quad (11)$$

where $\widetilde{\boldsymbol{M}} = \widetilde{\boldsymbol{m}}^{\mathsf{H}}\widetilde{\boldsymbol{m}} \in \mathbb{C}^{(L+1)\times(L+1)}$ and $\widetilde{\boldsymbol{U}} = \widetilde{\boldsymbol{u}}\widetilde{\boldsymbol{u}}^{\mathsf{H}}$. As in [20] and [12], problem (11) can be solved by means of a block coordinate descent algorithm, which we summarize in Algorithm 2 (lines 1 to 11). It consists of a nested loop where at each iteration i, the columns of $\widetilde{\boldsymbol{U}}$ are updated sequentially using the notation:

$$k^c = \{0, \ldots, k-1, k+1, \ldots, L-1\}. \quad (12)$$

This yields a global solution $\widetilde{\boldsymbol{U}}$ to problem (11). If this solution is of rank 1, i.e., $\widetilde{\boldsymbol{U}} = \widetilde{\boldsymbol{u}}\widetilde{\boldsymbol{u}}^{\mathsf{H}}$, then $\widetilde{\boldsymbol{u}}[0, \ldots, L-1]/\widetilde{\boldsymbol{u}}[L]$ is guaranteed to globally solve (8). We then obtain a global solution of (3) via:

$$\boldsymbol{x}_{\bar{v}} = \boldsymbol{\Phi}_{\bar{v}}^{\mathsf{H}} \operatorname{diag}(\boldsymbol{b})\widetilde{\boldsymbol{u}}[0, \ldots, L-1]/\widetilde{\boldsymbol{u}}[L]. \quad (13)$$

However, $\widetilde{\boldsymbol{U}}$ needs not be rank-1 in general. Hence, as commonly done in semi-definite relaxations, its closest rank-1 approximation is used in practice, which by the Eckart-Young-Mirsky theorem amounts to setting $\widetilde{\boldsymbol{u}}$ to the eigenvector associated to the largest eigenvalue of $\widetilde{\boldsymbol{U}}$.

Algorithm 2 CR for signal inpainting

Input: $\begin{cases} \boldsymbol{b} \in \mathbb{R}_+^L : \text{observations}, \ \boldsymbol{x}_v^\natural : \text{known signal} \\ \boldsymbol{\Phi} : \text{Fourier matrix}, \ \nu \geq 0 : \text{barrier parameter} \\ \widetilde{\boldsymbol{m}} \in \mathbb{C}^{L\times(L+1)} : \text{matrix defined by (9)} \\ n_{\text{iter}} : \text{number of iterations} \end{cases}$

1: $\widetilde{\boldsymbol{M}} \leftarrow \widetilde{\boldsymbol{m}}^{\mathsf{H}}\widetilde{\boldsymbol{m}}$ and $\widetilde{\boldsymbol{U}}^{(0)} \leftarrow \boldsymbol{I}$
2: **for** $i = 1, \ldots, n_{\text{iter}}$ **do**
3: **for** $k = 0, \ldots, L-1$ **do**
4: $\boldsymbol{z} \leftarrow \widetilde{\boldsymbol{U}}_{k^c,k^c}^{(i)}\widetilde{\boldsymbol{M}}_{k^c,k}$ and $\gamma \leftarrow \boldsymbol{z}^{\mathsf{H}}\widetilde{\boldsymbol{M}}_{k^c,k}$
5: **if** $\gamma > 0$ **then**
6: $\widetilde{\boldsymbol{U}}_{k^c,k}^{(i+1)}$ and $\left(\widetilde{\boldsymbol{U}}_{k^c,k}^{(i+1)}\right)^{\mathsf{H}} \leftarrow -\sqrt{\frac{1-\nu}{\gamma}}\boldsymbol{z}$
7: **else**
8: $\widetilde{\boldsymbol{U}}_{k^c,k}^{(i+1)}$ and $\left(\widetilde{\boldsymbol{U}}_{k^c,k}^{(i+1)}\right)^{\mathsf{H}} \leftarrow 0$
9: **end if**
10: **end for**
11: **end for**
12: $\widetilde{\boldsymbol{u}}$ = eigenvector associated to the largest eigenvalue of $\widetilde{\boldsymbol{U}}^{(i+1)}$, and $\widetilde{\boldsymbol{u}} \leftarrow \frac{\widetilde{\boldsymbol{u}}}{\widetilde{\boldsymbol{u}}[L-1]}$
13: $\boldsymbol{x}_{\bar{v}} \leftarrow \Re\left(\boldsymbol{\Phi}_{\bar{v}}^{\mathsf{H}} \operatorname{diag}(\boldsymbol{b})\widetilde{\boldsymbol{u}}[:L-1]\right)$ and $\boldsymbol{x} = \begin{bmatrix} \boldsymbol{x}_{\bar{v}} \\ \boldsymbol{x}_v^\natural \end{bmatrix}$

Output: Reconstructed signal $\boldsymbol{x}$

III. Experiments

In this section we assess the potential of our methods for inpainting gaps in audio signals. Our code is available online.[3]

A. Experimental setting

a) Data: We consider 100 speech signals from the Librispeech dataset [29]. Signals are sampled at 16 kHz. For each signal, we extract a non-silent sub-signal of variable length L at a random location, where we create a gap of d samples.

b) Proposed methods: For the AM algorithm, iterations stop when the the maximum loss variation over the previous 5 iterations does not exceed 10^{-10}, or when a maximum number of 1000 iterations is reached. Preliminary experiments revealed no significant difference between various basic initialization schemes, e.g., using a random or zero phase in the Fourier domain, or using a random or zero missing signal in the time domain. We also derived an initialization scheme inspired by the spectral initialization method in [30], but this did not yield any significant improvement either. Therefore, the results displayed hereafter use $\boldsymbol{x}_{\bar{v}}^{(0)} \leftarrow \mathbf{0}$. The CR algorithm uses a fixed amount $n_{\text{iter}} = 10$ since the performance did not show further improvement beyond in our experiments. We simply set the barrier parameter ν to 0 as in [12]. We also consider a *combined* CR+AM algorithm. This technique consists in first estimating the signal with CR, and then using this estimate as an initialization for AM, with the same stopping criterion as above. Our methods are fed with the ground truth magnitudes ($\boldsymbol{b} = |\boldsymbol{\Phi}\boldsymbol{x}^\natural|$), except in the last experiment where noisy magnitudes are considered.

[3] https://github.com/Louis-Bahrman/Inpainting-Fourier

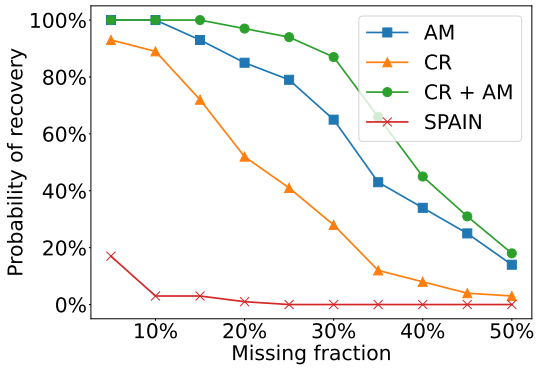

Fig. 2. Influence of the missing fraction onto performance.

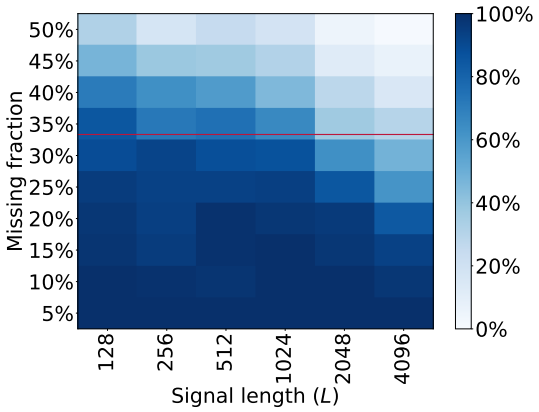

Fig. 3. Probability of achieving perfect reconstruction for CR+AM, as a function of the signal length (in samples) and the missing signal fraction. The red line represents the 33 % theoretical limit under which perfect reconstruction is achievable for almost all signals [23].

c) Baselines: Let us note that typical phase retrieval methods [17], [20] are not appropriate comparison baselines since they are agnostic to the observed samples. Besides, more recent phase retrieval approaches proposed in audio [22], [31] are tailored to time-frequency processing, and are therefore not suitable for the setup considered in this paper. On the other hand, traditional audio restoration techniques such as [6] do not leverage magnitude knowledge, making comparisons unfair. Nevertheless, results obtained using the Fourier-sparsity-based method SPAIN [8] are shown, as an indication of the potential of exploiting magnitude models beyond sparsity. The same parameters as in [8] are used, except for the relaxation parameter r which is set to 32.

d) Metric: To assess the quality of the reconstruction, we resort to the signal-to-error ratio (SER) expressed in dB:

$$\mathrm{SER}(\boldsymbol{x}_{\bar{v}}, \boldsymbol{x}^{\natural}_{\bar{v}}) = 10\log_{10}\left(\frac{\left\|\boldsymbol{x}^{\natural}_{\bar{v}}\right\|^2}{\left\|\boldsymbol{x}_{\bar{v}} - \boldsymbol{x}^{\natural}_{\bar{v}}\right\|^2}\right), \qquad (14)$$

where $\boldsymbol{x}$ denotes the estimated signal (higher is better). Let us outline that the SER is only calculated over the set of missing samples $\bar{v}$ since $\boldsymbol{x}_v = \boldsymbol{x}^{\natural}_v$ elsewhere. Note that AM's initialization (replacing the missing signal with zeros) corresponds to an SER of 0. Based on preliminary listening tests, we consider that perfect reconstruction is achieved when the SER is greater than 20 dB.

B. Results

First, we compare the recovery rate of the three proposed methods with respect to the missing signal fraction. The results corresponding to $L = 1024$ (64 ms) are displayed in Fig. 2. Our methods significantly outperform SPAIN in this setting, regardless of the missing fraction. This demonstrates the potential of exploiting Fourier magnitudes when these are available. CR+AM achieves the best results and consistently outperforms the other approaches. This is explained by the ability of CR to provide a solution that is more likely to converge to a global minimum than AM's basic initialization. While all the methods exhibit a performance drop when the missing fraction increases, CR+AM and AM still yield perfect reconstruction in at least 80 % of the cases when less than 30 % of the signal is missing. The results of CR+AM have to be weighted against the higher memory and time complexity of CR, which processes matrices of size $L \times L$ instead of vectors of size L. Accordingly, a general guideline might be to resort to the AM method when very few samples are missing (5 to 10 %), and to turn to CR+AM in all remaining scenarios, at a higher computational cost.

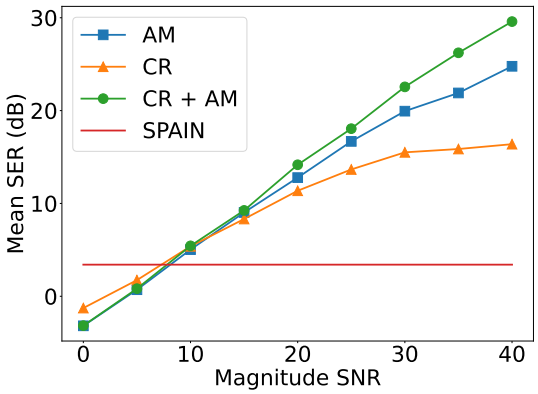

Fig. 4. Influence of the magnitude noise onto performance.

Then, Fig. 3 shows the recovery rate achieved by the best performing method CR+AM as a function of the signal length L (which ranges from 128 to 4096 samples) and the fraction of missing samples d/L (which ranges from 5 % to 50 %). We observe that perfect reconstruction is achieved in most cases when the fraction is lower than 33 % and $L \leq 1024$, suggesting that the approach performs near the theoretical optimum of [23] in this regime. While for short signals the method even yields satisfactory results under slightly larger missing fractions, its performance eventually drops for larger signals. This may be explained by the increased dimensionality of the problem which may increase the risk of getting trapped in local minima.

Finally, let us evaluate the robustness of our methods to magnitude noise. Indeed, in the previous experiments we considered oracle magnitudes, but in practical applications these are estimated beforehand and therefore contaminated by

estimation errors. To simulate such a scenario, we consider noisy magnitudes $\boldsymbol{b} = \max\left(0, \left|\boldsymbol{\Phi}\boldsymbol{x}^\natural\right| + \boldsymbol{n}\right)$, where $\boldsymbol{n}$ is a white Gaussian noise whose variance is adjusted to fit a given magnitude signal-to-noise (SNR) value. The results are presented in Fig. 4, where the signal length is fixed at $L = 1024$ and the fraction of missing signal at 25%. Overall, we observe that all of our methods exhibit a similar linear decay in performance in the log-log plot when the SNR falls below 20 dB, suggesting an encouraging robustness to errors. Interestingly, while CR performs worse than the other methods at higher SNRs, it still yields solutions that provide a better initialization than zero to AM. Finally, we remark that SPAIN, which leverages Fourier-domain sparsity, outperforms our methods at low SNRs, while the opposite is true for SNRs above 8 dB. This confirms that leveraging Fourier magnitude is promising, provided they have been accurately estimated beforehand. This motivates future research on magnitude estimation techniques.

IV. Conclusion

We have investigated the problem of signal inpainting from observed Fourier magnitude measurements. After formulating the associated optimization problem, we proposed two estimation methods based on alternating minimization and convex relaxation. Experiments in the noiseless regime highlighted that combining the two methods yields near-exact reconstruction most of the time when less than 33% of the signal is missing, closely following our theoretical limit. Further experiments demonstrated a relative robustness of the methods to magnitude errors. Future work will therefore focus on applying such techniques to more realistic settings where magnitudes are estimated, e.g., via light magnitude interpolation methods in the short-time Fourier domain or via data-driven models such as deep neural networks.

References

[1] A. Adler, V. Emiya, M. G. Jafari, M. Elad, R. Gribonval, and M. D. Plumbley, "Audio inpainting," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 3, pp. 922–932, 2012.

[2] C. Rodbro, M. Murthi, S. Andersen, and S. Jensen, "Hidden Markov model-based packet loss concealment for voice over IP," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 5, pp. 1609 –1623, September 2006.

[3] G. Chantas, S. Nikolopoulos, and I. Kompatsiaris, "Sparse audio inpainting with variational bayesian inference," in *Proc. IEEE International Conference on Consumer Electronics (ICCE)*, October 2018.

[4] D. Derebssa, K. Eneman, and T. Van Waterschoot, "Missing sample estimation based on high-order sparse linear prediction for audio signals," in *Proc. European Signal Processing Conference (EUSIPCO)*, September 2018.

[5] P. Magron, R. Badeau, and B. David, "Phase reconstruction of spectrograms with linear unwrapping: Application to audio signal restoration," in *Proc. European Signal Processing Conference (EUSIPCO)*, August 2015.

[6] A. Janssen, R. Veldhuis, and L. Vries, "Adaptive interpolation of discrete-time signals that can be modeled as autoregressive processes," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 34, no. 2, pp. 317–330, 1986.

[7] G. Taubock, S. Rajbamshi, and P. Balazs, "Dictionary learning for sparse audio inpainting," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 1, pp. 104–119, January 2021.

[8] O. Mokry, P. Zaviska, P. Rajmic, and V. Vesely, "Introducing SPAIN (SParse audio INpainter)," in *Proc. European Signal Processing Conference (EUSIPCO)*, September 2019.

[9] S. Godsill and P. Rayner, "A bayesian approach to the restoration of degraded audio signals," *IEEE Transactions on Speech and Audio Processing*, vol. 3, no. 4, pp. 267–278, 1995.

[10] M. Lagrange, S. Marchand, and J.-B. Rault, "Long interpolation of audio signals using linear prediction in sinusoidal modeling," *Journal of the Audio Engineering Society*, vol. 53, no. 10, pp. 891–905, 2005.

[11] N. Perraudin, N. Holighaus, P. Majdak, and P. Balazs, "Inpainting of long audio segments with similarity graphs," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 6, pp. 1083–1094, 2018.

[12] A. Deleforge and Y. Traonmilin, "Phase unmixing: Multichannel source separation with magnitude constraints," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017.

[13] A. Liutkus, C. Rohlfing, and A. Deleforge, "Audio source separation with magnitude priors: the BEADS model," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2018.

[14] C. Févotte, N. Bertin, and J.-L. Durrieu, "Nonnegative matrix factorization with the Itakura-Saito divergence: With application to music analysis," *Neural computation*, vol. 21, no. 3, pp. 793–830, March 2009.

[15] L. Girin, F. Roche, T. Hueber, and S. Leglaive, "Notes on the use of variational autoencoders for speech and audio spectrogram modeling," in *Proc. International Conference on Digital Audio Effects (DAFx)*, 2019.

[16] A. Walther, "The question of phase retrieval in optics," *Optica Acta: International Journal of Optics*, vol. 10, no. 1, pp. 41–49, 1963.

[17] R. W. Gerchberg and O. W. Saxton, "A practical algorithm for the determination of phase from image and diffraction plane pictures," *Optik*, vol. 35, p. 237–246, 1972.

[18] E. J. Candès, X. Li, and M. Soltanolkotabi, "Phase retrieval via Wirtinger flow: Theory and algorithms," *IEEE Transactions on Information Theory*, vol. 61, no. 4, pp. 1985–2007, April 2015.

[19] E. J. Candès, T. Strohmer, and V. Voroninski, "Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming," *Communications on Pure and Applied Mathematics*, vol. 66, no. 8, pp. 1241–1274, 2013.

[20] I. Waldspurger, A. D'Aspremont, and S. Mallat, "Phase recovery, maxcut and complex semidefinite programming," *Mathematical Programming*, vol. 149, no. 1, pp. 47–81, 2015.

[21] Z. Li, K. Lange, and J. A. Fessler, "Poisson phase retrieval with Wirtinger flow," in *Proc. IEEE International Conference on Image Processing (ICIP)*, September 2021.

[22] P.-H. Vial, P. Magron, T. Oberlin, and C. Févotte, "Phase retrieval with Bregman divergences and application to audio signal recovery," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 1, pp. 51–64, January 2021.

[23] M. Krémé, A. Deleforge, P. Magron, and L. Bahrman, "Signal inpainting from Fourier magnitudes: An almost uniqueness result," https://magronp.github.io/files/2023inpainting_sup.pdf, February 2023.

[24] T. Bendory, R. Beinert, and Y. C. Eldar, *Fourier Phase Retrieval: Uniqueness and Algorithms*. Springer, 2017, pp. 55–91.

[25] L. Xu, P. Yan, and T. Chang, "Almost unique specification of discrete finite length signal: from its end point and Fourier transform magnitude," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, April 1987.

[26] S. H. Nawab, T. F. Quatieri, and J. S. Lim, "Signal reconstruction from short-time Fourier transform magnitude," in *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 31, August 1983, pp. 986–998.

[27] C. Audet, P. Hansen, B. Jaumard, and G. Savard, "A branch and cut algorithm for nonconvex quadratically constrained quadratic programming," *Mathematical Programming*, vol. 87, no. 1, pp. 131–152, 2000.

[28] Y. G. Yevtushenko and V. G. Zhadan, "Exact auxiliary functions in optimization problems," *USSR computational mathematics and mathematical physics*, vol. 30, no. 1, pp. 31–42, 1990.

[29] V. Panayotov, G. Chen, D. Povey, and S. Khudanpur, "Librispeech: An ASR corpus based on public domain audio books," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2015.

[30] E. J. Candès, X. Li, and M. Soltanolkotabi, "Phase retrieval via Wirtinger flow: Theory and algorithms," *IEEE Transactions on Information Theory*, vol. 61, no. 4, pp. 1985–2007, April 2015.

[31] Y. Masuyama, K. Yatabe, and Y. Oikawa, "Griffin–Lim like phase recovery via alternating direction method of multipliers," *IEEE Signal Processing Letters*, vol. 26, no. 1, pp. 184–188, January 2019.